AD-A285 448

Journal Offprint Paper

Accesion For		
NTIS CRA&I		X
DTIC TAB		☐
Unannounced		☐
Justification		
By		
Distribution /		
Availability Codes		
Dist	Avail and / or Special	
A-1	20	

BUTTERWORTH
HEINEMANN

A member of the Reed Elsevier group

Linacre House, Jordan Hill, Oxford OX2 8DP, UK

Best Available Copy

OFFPRINT PAPERS

How normal flow constrains relative depth for an active observer

Liuqing Huang and Yiannis Aloimonos

We present a set of constraints that relate the relative depth of (stationary or moving) objects in the field of view with the spatiotemporal derivatives of the time varying image intensity function. The constraints are purposive in the sense that they can be used only for the relative depth from motion problem and not in other problems related to motion (i.e. they lack generality). In addition, they show that relative depth could be obtained without having to go through the intermediate step of fullly recovering 3D motion, as is commonly considered. Our analysis indicates that exact computation of retinal motion (optic flow or displacements) does not appear to be a necessary first step for some problems related to visual motion, contrary to conventional wisdom. In addition, it is demonstrated that optic flow, whose computation is an ill-posed problem, is related to the motion of the scene only under very restrictive assumptions. This paper is devoted to the discovery of the mathematical constraints relating normal flow and relative depth. The development of algorithms using these constraints and the study of stability issues of such algorithms, is not discussed here.

Keywords: computer vision, constraints, field of view

The problem of structure from motion has attracted a lot of attention in the past few years[1-14] because of the general usefulness that a potential solution to this problem would have. Important navigational problems such as detection of independently moving objects by a moving observer, passive navigation, obstacle detection, target pursuit and many other problems related to robotics, teleconferencing, etc. would be simple applications of a structure from motion module. The problem has been formulated as follows: Given a sequence of images taken by a monocular observer (the observer and/or parts of the scene could be moving), to recover the shapes (and relative depths) of the objects in the scene, as well as the (relative) 3D motions of independently moving bodies.

The problem has been formulated and usually treated as an aspect of the general task of recovering 3D information from motion[15,16]. The majority of the proposed solutions to date are based on the following modular approach:

1. First, one computes the optic flow on the image plane, i.e. the velocity with which every image point appears to be moving. (For clarity, we consider only the differential case. In the case of long range motion one computes discrete displacements, but the analysis remains essentially the same.)
2. Then segmentation of the flow field is performed and different moving objects are identified on the image plane. From the segmented optic flow one then computes the 3D motion with which each visible surface is moving relative to the observer. (Assuming that an object moves rigidly, a monocular observer can only compute its direction of translation and its rotation, but not its speed).
3. Finally, using the values of the optic flow, along with the results of the previous step, one computes the surface normal at each point, or equivalently, the ratio Z_i/Z_j of the depths of any two points i and j.

The reason that most approaches have followed the above three-step approach is two-fold. The first is due to the formulation of the problem, which insists on recovering a complete relative depth map and accurate three-dimensional motion. The second is due to the fact that the constraints relating retinal motion to three-

Computer Vision Laboratory, Center for Automation Research, University of Maryland, College Park, MD 20742-3275, USA
Paper received: 8 March 1993; revised paper received: 9 February 1994

0262-8856/94/07/0435-11

dimensional structure involve 3D motion in a nonlinear manner that does not allow separability. For examples of such approaches, see elsewhere[1–6,17–19]. However, the past work in this paradigm, despite its mathematical elegance, is far from being useful in real-time navigational systems, and such techniques have found few or no practical applications (possible exceptions are photogrammetry and semiautonomous applications requiring a human operator). Consequently, this approach cannot be used to explain the ability of biological organisms to handle visual motion.

There exist many reasons for the limitations of the optic flow approach, related to all three steps listed above. To begin, the computation of optic flow is an ill-posed problem, i.e. unless we impose additional constraints, we cannot estimate it[20]. Such constraints, however, impose a relationship on the values of the flow field which is translated into an assumption about the scene in view (for example, smooth). Thus, even if we are capable of obtaining an algorithm that computes optic flow in a robust manner, the algorithm will work only for a restricted set of scenes. The only available constraint at every point (x, y) of the changing image $I(x, y, t)$ for the flow (u, v) is the constraint $I_x u + I_y v + I_t = 0$[21], where the subscripts denote partial differentiation. This means that we can only compute the projection of the flow on the gradient direction $((I_x, I_y)\cdot(u, v) = -I_t)$, i.e. the so-called normal flow. More graphically, it means that if a feature (for example, an edge segment) in the image moves to a new position, we don't know where every point of the segment moved to (see *Figure 1*); we only know the normal flow, i.e. the projection of the flow on the image gradient at that point.

A second reason has to do with the very essence of optic flow. An optic flow field is the vector field of apparent velocities that are associated with the variation of brightness on the image plane. Clearly, the scene is not involved in this definition. One would hope that optic flow is equivalent to the so-called motion field[16], which is the (perspective) projection on the image plane of the three-dimensional velocity field associated with each point of the visible surfaces in the scene. However, the optic flow field and the motion field are not equal in general. Verri and Poggio[22] reported some general results in an attempt to quantify the difference between the optic flow and motion fields. Although we don't yet have necessary and sufficient conditions for the equality of the two fields, it is clear that they are equal only under specific sets of restrictive conditions.

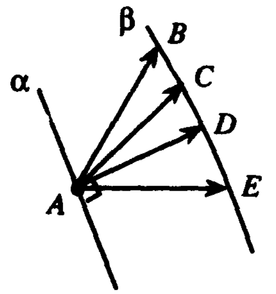


Figure 1 **The aperture problem. Point *A* could have moved to *B*, *C*, *D*, *E*. However, whatever the value of the image motion vector is, its projection on the normal to α is always *AD* (known)**

A third reason is related to the second step of the existing algorithms for structure from motion. These algorithms attempt to first recover three-dimensional motion before they proceed to recover relative depth, and this problem of 3D motion appears to be very sensitive in the presence of small amounts of noise in the input (flow or displacements\hbox{()})[17,19,23,24].

Is it possible to compute relative depth from motion without using optic flow fields (which are difficult to compute and in general not equal to the motion fields), and without having to go through the intermediate stage of 3D motion recovery? If it is, then we have the potential for a more robust algorithm. This is the question we study in this paper. It turns out that it is indeed possible to compute relative depth if we use the spatiotemporal derivatives of the image intensity function and we employ an active observer.

INPUT

Our motivation is by now clear. We wish to avoid using optic flow as the input to the computation of structure from motion. On the other hand, we must utilize some description of the image motion. As such a description we choose the spatial and temporal derivatives $\frac{\partial I}{\partial x}, \frac{\partial I}{\partial y}, \frac{\partial I}{\partial t}$, of the image intensity function $I(x, y, t)$. These quantities define the normal flow at every point, i.e. the projection of the optic flow on the direction of the gradient (I_x, I_y). Clearly, estimating the normal flow is much easier than estimating the actual optic flow. But how is normal flow related to the three-dimensional motion field? Is the normal optic flow field equal to the normal motion field, and under what conditions? This question was addressed by Verri and Poggio[2].

Let $I(x, y, t)$ denote the image intensity, and consider the optic flow field $\vec{v} = (u, v)$ and the motion field $\vec{\bar{v}} = (\bar{u}, \bar{v})$ at a point (x, y) where the local (normalized) intensity gradient is $\vec{n} = (I_x, I_y)/\sqrt{I_x^2 + I_y^2}$. The normal motion field at point (x, y) is by definition:

$$\bar{u}_n = \vec{\bar{v}}\cdot\vec{n} \qquad \text{or}$$

$$\bar{u}_n = \frac{(I_x, I_y)}{\sqrt{I_x^2 + I_y^2}}\cdot\left(\frac{dx}{dt}, \frac{dy}{dt}\right) \qquad \text{or}$$

$$\bar{u}_n = \frac{\nabla I}{\|\nabla I\|}\cdot\left(\frac{dx}{dt}, \frac{dy}{dt}\right) \qquad \text{or}$$

$$\bar{u}_n = \frac{1}{\|\nabla I\|}\cdot\left(I_x\frac{dx}{dt} + I_y\, dy\right)$$

Similarly, the normal-optic flow[21] is:

$$u_n = -\frac{1}{\nabla I} I_t$$

Thus:

$$\tilde{u}_n - u_n = \frac{1}{\nabla I}\frac{dI}{dt}$$

From this equation it follows that if the change of intensity of an image patch during its motion $\left(\frac{dI}{dt}\right)$ is small enough (which is a reasonable assumption) and the local intensity gradient has a high magnitude, then the normal optic flow and motion fields are approximately equal. Thus, provided that we measure normal flow in regions of high local intensity gradients, the normal flow measurements can safely be used for inferring 3D structure.

PREVIOUS WORK AND PURPOSIVE VISION

The idea of using the spatiotemporal derivatives of the image intensity function for the solution of the structure from motion problem is not new. (Working with normal flow or the derivatives of the image is exactly the same thing. The difference is that the use of normal flow provides geometric intuition.) In Aloimonos and Brown[25] the case of rotational motion was examined. In Horn and Weldon[26] and Negahdaripour[27] the case of translational motion was examined in detail. Elsewhere, the general case was examined for recovering only 3D motion[28–32], using pattern matching.

In this paper, we take a purposive approach[33]. We would like to compute relative depth from motion without having to go through the estimation of 3D motion and without having to compute optic flow. In simple words, we want a procedure that computes relative depth and is designed only for this problem. Of course, if information about 3D motion is known, it can be effectively utilized in our problem, but this is of no concern to us here. When building a system that can deal with visual motion problems, we can visualize it as consisting of many processes working in a cooperative manner to solve various problems. For example, the theory described in this paper could be used to design a process that computes relative depth from image measurements, independently of the process that computes 3D motion. However, after a number of computational steps, when results about relative depth and 3D motion become available from the two independent processes, they can be exchanged and the constraints relating to them can be effectively utilized so that the results are as consistent as possible. Such an approach to building vision systems is less modular than the general recovery approach[15].

This approach of attempting general solutions to specific problems (purposive vision), as opposed to working towards solutions to general problems (reconstructionist vision), is justified by the potential robustness of the proposed solutions, and is very much needed for the development of successful systems in the real world. Of course, normal flow contains much less information than optical flow, and we cannot expect that we will be able to fully recover the relative depth map. Indeed, we show that for the case of moving objects, relative depth cannot be obtained everywhere (i.e. at every pixel), but only at points where the local intensity gradient is parallel to a given direction. But a full depth map is not always required. We only need the values of the depth that are relevant to the task at hand.

PAPER ORGANIZATION

We define the relative depth from motion problem as follows: 'Given an active observer that can collect a series of images of a scene, to recover the relative depths of objects (or features) in the scene.' (An *active observer*[20] controls the geometric parameters of its sensory apparatus, thus introducing constraints on its sensory data.)

Since the input to the perceptual process is the normal flow, and the normal flow field contains, in general, less information than the motion field, to solve the problem we need to transfer much of the computation to the activity of the observer[20]. A geometric model of the observer is given in *Figure 2*. Notice that the camera is resting on a platform ('neck') with six degrees of freedom (actually only one of the degrees is used), and the camera can rotate around its x and y axes (saccades). (However, in this work the only activity required is acceleration along the optical axis.)

The organization of the paper reflects the increasing difficulty of the problem as the motion of the object in view becomes more complex. The following section is devoted to the case of stationary objects. It is shown that if the observer moves along its optical axis, relative depth is easily obtained from the normal flow. Then we study the problem for the case of an object translating parallel to the image plane, deal with the case where the object is moving with a general translation, and analyse the general case. We assume that independently moving objects can be detected and localized on the image. This

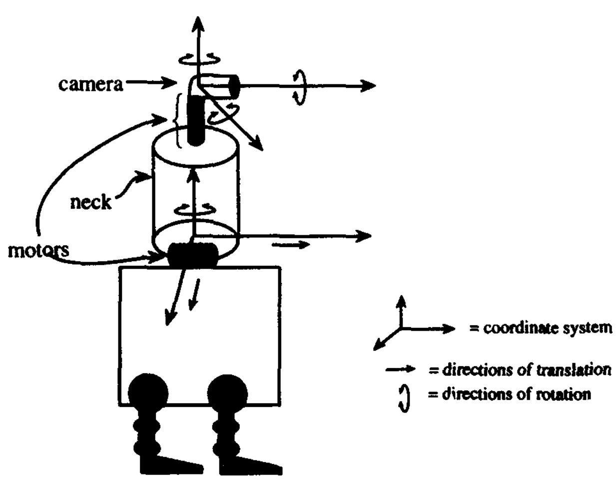


Figure 2 The active observer

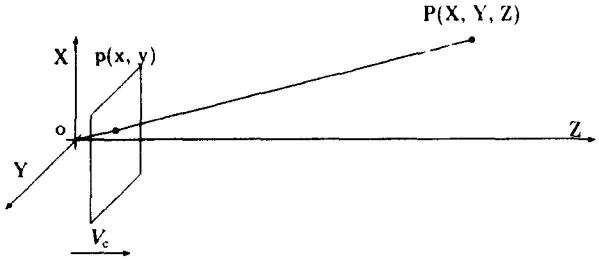


Figure 3 The camera moves towards the objects in a scene with velocity V_c

problem, which is nontrivial if the observer is moving, is addressed elsewhere[34-36].

STATIONARY OBJECTS

Let the camera move towards the scene with velocity V_c along its optical axis. Let the image point $p(x, y)$ be the projection of 3D point $P(X, Y, Z)$. After time dt, $P'(X, Y, Z - V_c dt)$, which is the new position of P, projects to $p'(x', y')$. Using the relations of perspective projection assuming unit focal length, we have (*Figure 3*):

$$x = \frac{X}{Z} \tag{1}$$

$$y = \frac{Y}{Z} \tag{2}$$

$$x' = \frac{X}{Z - V_c dt} \tag{3}$$

$$y' = \frac{Y}{Z - V_c dt} \tag{4}$$

Thus we can obtain the motion velocity of image point $p(x, y)$ as:

$$v_x \lim_{dt \to 0} \frac{x' - x}{dt} = \frac{V_x}{Z} + x\frac{V_c}{Z} \tag{5}$$

Similarly, we have:

$$v_y \lim_{dt \to 0} \frac{y' - y}{dt} = \frac{V_y}{Z} + y\frac{V_c}{Z} \tag{6}$$

Suppose the unit normal vector (i.e. the direction of the image gradient) $p(x, y)$ is (n_x, n_y). The normal vector is the projection of the motion field on the unit normal vector. Thus we have the following relationship between the motion velocity and the normal flow:

$$v_n = v_x n_x + v_y n_y \tag{7}$$

or:

$$v_n = \frac{V_c}{Z} x n_x + \frac{V_c}{Z} y n_y$$

$$= \frac{V_c}{Z}(x n_x + y n_y)$$

or:

$$Q = \frac{V_c}{Z} = \frac{v_n}{x n_x + y n_y} \tag{8}$$

As the camera is the only moving object in the scene, and all the objects are stationary, V_c is the same for all image points. Thus we can use equation (8) to decide which object or feature is closer.

OBJECT TRANSLATING PARALLEL TO THE FOCAL PLANE

Here we study the case where the object is translating parallel to the focal plane with velocity V_x, V_y, along the x and y axes respectively, while the camera is moving towards the object with velocity V_c along the z axis. The velocity of the object with respect to the camera is $(V_x, V_y, -V_c)$. Assume that point $P(X, Y, Z)$ projects to $p(x, y)$ at time t, and after time dt the same point $P(X + V_x dt, Y + V_y dt, Z - V_c dt)$ projects to $p(x', y')$; then we have (see *Figure 4*):

$$x = \frac{X}{Z} \tag{9}$$

$$y = \frac{Y}{Z} \tag{10}$$

$$x' = \frac{X + V_x dt}{Z - V_c dt} \tag{11}$$

$$y' = \frac{Y + V_y dt}{Z - V_c dt} \tag{12}$$

Thus we can obtain the motion velocity of image point $p(x, y)$ as:

$$v_x \lim_{dt \to 0} \frac{x' - x}{dt} = \frac{V_x}{Z} + x\frac{V_c}{Z} \tag{13}$$

Similarly, we have:

$$v_y \lim_{dt \to 0} \frac{y' - y}{dt} = \frac{V_y}{Z} + y\frac{V_c}{Z} \tag{14}$$

According to equation (7) we have:

$$v_n = \frac{V_c}{Z}(x n_x + y n_y) + \frac{V_c}{Z}\left(n_x \frac{V_x}{V_c} + n_y \frac{V_y}{V_c}\right) \tag{15}$$

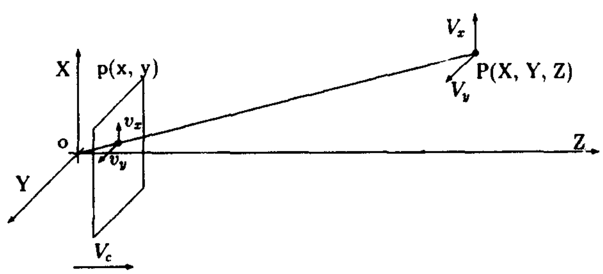


Figure 4 The object is moving parallel to the focal plane

While we cannot immediately recover $(V_x/V_c, V_y/V_c)$ from the images, the vector is parallel to the direction of motion of the object on the xy-plane (V_x, V_y). In the Appendix we show how to estimate the direction of (V_x, V_y) (i.e. V_x/V_y) in the general case. Note that in natural scenes of objects, normal flows are available in all directions. If we select a normal vector from the image of the object that is perpendicular to the direction of motion, the second term of equation (15) will be zero. Thus for objects moving parallel to the focal plane, we obtain the direction of motion (V_x, V_y) (see Appendix). Then, for normal flows that are perpendicular to the direction of motion, we have:

$$Q = \frac{V_c}{Z} = \frac{v_n}{xn_x + yn_y} \tag{16}$$

It is noteworthy that partial 3D motion information (V_x/V_y) is utilized in this case.

OBJECT WITH GENERAL TRANSLATION

When an object is translating with velocity (V_x, V_y, V_z) with respect to the camera while the camera is translating along the z axis with velocity V_c, the motion of the object with respect to the coordinate system centred at the camera is $(V_x, V_y, V_z - V_c)$ (*Figure 5*). According to equation (16), if we select normal flows perpendicular to the direction of motion, we have:

$$\frac{V_c - V_z}{Z} = \frac{v_n}{xn_x + yn_y} \tag{17}$$

This measurement is not useful yet because we have an object-specific velocity V_z.

To eliminate the unknown V_z, the translational velocity of the moving object along the z axis, we will use two consecutive frames, at times t_1 and t_2. Assume that the scene consists of a stationary and a moving object: that the stationary object at time t_1 is at $P(X_{11}, Y_{11}, Z_{11})$, and at time t_2 is at $P(X_{12}, Y_{12}, Z_{12})$; and that the moving object at time t_1 is at $P(X_{21}, Y_{21}, Z_{21})$, and at time t_2 is at $P(X_{22}, Y_{22}, Z_{22})$. We also assume that the velocity of the camera at time t_1 is V_c and at time t_2 is cV_c, where $c \neq 1$ is a constant. If the camera is accelerating much faster than the object, we can assume that the velocity of the object remains the same across the frames. We select a normal flow v_n that is perpendicular to the direction of the 3D motion in the xy plane.

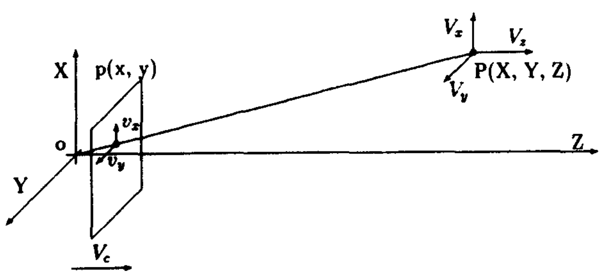


Figure 5 Moving robot hand approaching a stationary object

From equation (16) we have:

$$\frac{V_c}{Z_{11}} = b_{11} \tag{18}$$

$$\frac{cV_c}{Z_{12}} = b_{12} \tag{19}$$

$$\frac{V_c - V_z}{Z_{21}} = b_{21} \tag{20}$$

$$\frac{cV_c - V_z}{Z_{22}} = b_{22} \tag{21}$$

and:

$$Z_{12} = Z_{11} - V_c dt \tag{22}$$

$$Z_{22} = Z_{21} - (V_c - V_z)dt \tag{23}$$

From the above equations, when dt is small, we obtain:

$$\frac{V_c}{Z_{12}} = \frac{b_{11} - b_{12} + b_{11}b_{12}dt}{1 - c - b_{11}dt + cb_{11}dt} = \frac{b_{11} - b_{12}}{1 - c} \tag{24}$$

and:

$$\frac{V_c}{Z_{22}} = \frac{b_{21} - b_{22} + b_{21}b_{22}dt}{1 - c - b_{21}dt + cb_{21}dt} = \frac{b_{21} - b_{22}}{1 - c} \tag{25}$$

or:

$$Q(Z_{12}, V_c, 1 - c) = \frac{V_c(1 - c)}{Z_{12}} = b_{11} - b_{12} \tag{26}$$

and:

$$Q(Z_{22}, V_c, 1 - c) = \frac{V_c(1 - c)}{Z_{22}} = b_{21} - b_{22} \tag{27}$$

where for $i, j = 1, 2$:

$$b_{ij} = \frac{u_{nij}}{x_{ij}n_{xij} + n_{ij}n_{yij}} \tag{28}$$

Thus we have obtained the relative depth function Q for a moving object and a stationary object. Velocity V_c and velocity ratio c are not known, but since they are parameters of the camera, they remain the same for all objects involved. We assume that it is known whether the camera is moving forward or backward; thus we know the sign of V_c. We also assume that it is known whether the camera is accelerating or decelerating; thus we know the sign of $1 - c$. Therefore, we can determine the relative depth of the two objects from equation (28). (It is worth noting that the same results can be achieved if the camera is at first stationary and then moves quickly to a new position instead of moving and then accelerating. In this case, $V_c \to 0$, $c \to \infty$ and $c \cdot V_c \to V'_c$. Thus the relative measures become $Q(z, V'_c) = -\frac{V'_c}{Z}$).

OBJECT MOVING IN AN UNRESTRICTED RIGID MANNER

The motion of a rigid object can be described as the sum of a rotation plus a translation. We can choose a point through which the rotation axis passes; this gives a unique rotation and translation describing the rigid motion (in general, there are infinitely many combinations of rotations and translations describing the same rigid motion). Assume that the object is translating with velocity $T = (T_x, T_y, T_z)^T$ and rotating with angular velocity $R = (R_x, R_y, R_z)^T$ around a point $P = (X_0, Y_0, Z_0)$ on its surface (*Figure 6*). The translational components are measured with respect to the world coordinate system, while the angular velocity is measured with respect to the coordinate system whose origin is located at point $(X_0, Y_0, Z_0)^T$. The camera is moving with velocity T_c along the z axis.

Point P is visible in the image; its image is point $p = (x_0, y_0)$. We attach a coordinate system to the object, at point P, with axes parallel to the axes of the observer coordinate system. We express the motion of the object in this object-based coordinate system. The camera is moving with velocity T_c along the Z-axis. Then the velocity of any point Q on the object is:

$$\mathbf{V} = \begin{bmatrix} T_x \\ T_y \\ T_z - T_c \end{bmatrix} + R \times \begin{bmatrix} X - X_0 \\ Y - Y_0 \\ Z - Z_0 \end{bmatrix}$$

$$= \begin{bmatrix} T_x + R_y(Z - Z_0) - R_z(Y - Y_0) \\ T_y + R_z(X - X_0) - R_x(Z - Z_0) \\ T_z - T_c + R_x(Y - Y_0) - R_y(X - X_0) \end{bmatrix}$$

Thus, expressing the optic flow (v_x, v_y) on the image, we have:

$$\begin{aligned} v_x &= \frac{d}{dt}\frac{X}{Z} \\ &= \frac{V_x}{Z} - x\frac{V_c}{Z} \\ &= \frac{T_x}{Z} + x\frac{T_c - T_z}{Z} \\ &\quad - x(y - y_0)R_x + \left(\frac{Z - Z_0}{X} + x(x - x_0)\right) \\ &\quad R_y - (y - y_0)R_z \end{aligned}$$

and:

$$\begin{aligned} v_y &= \frac{d}{dt}\frac{Y}{Z} \\ &= \frac{V_y}{Z} - y\frac{V_c}{Z} \\ &= \frac{T_y}{Z} + y\frac{T_c - T_z}{Z} \\ &\quad - \left(\frac{Z - Z_0}{Z} + y(y - y_0)\right)R_x + y(x - x_0) \\ &\quad R_y + (x - x_0)R_z \end{aligned}$$

where (x_0, y_0) is the projection of $(X_0, Y_0, Z_0)^T$. Combining the above equations with equation (7), we obtain for the normal flow:

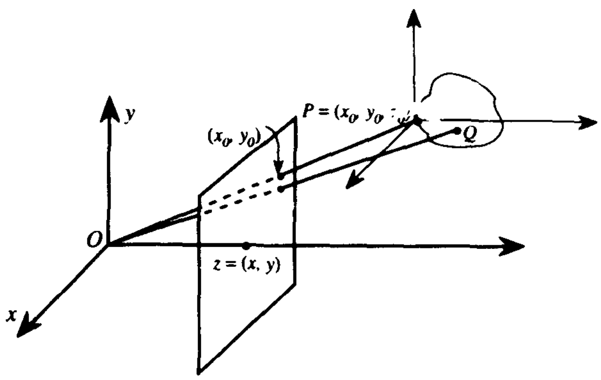


Figure 6 Object moving in an unrestricted rigid manner

$$\begin{aligned} v_n &= \frac{T_c - T_z}{Z}\left(n_x\frac{T_x}{T_c - T_z} + n_y\frac{T_y}{T_c - T_z}\right) \\ &\quad + \frac{T_c - T_z}{Z}(n_x x + n_y y) \\ &\quad - R_x(y - y_0)(xn_x + yn_y) \\ &\quad + R_y(x - x_0)(xn_x + yn_y) \\ &\quad + R_z((x - x_0)n_y - (y - y_0)n_x) \\ &\quad + \frac{Z - Z_0}{Z}(R_y n_x - R_x n_y) \end{aligned}$$

Considering this measurement u_n at point $x = x_0$, $y = y_0$ (and $Z = Z_0$) we have:

$$\begin{aligned} u_n &= \frac{T_c - T_z}{Z}\left(n_x\frac{T_x}{T_c - T_z} + n_y\frac{T_y}{T_c - T_z}\right) \\ &\quad + \frac{T_c - T_z}{Z}(xn_x + yn_y) \end{aligned}$$

Provided that the direction (n_x, n_y) of the normal flow at (x_0, y_0) is perpendicular to the direction of parallel translation (T_x, T_y), we get:

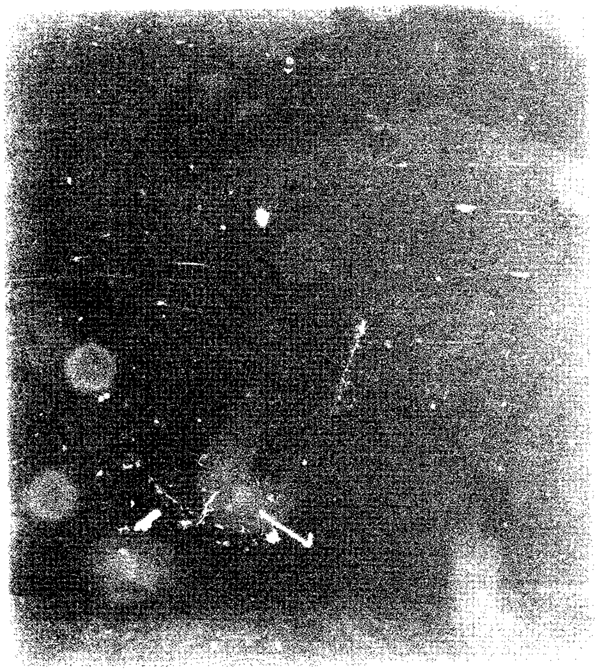
$$\frac{T_c - T_z}{Z} = \frac{u_n}{x_0 n_x + y_0 n_y}$$

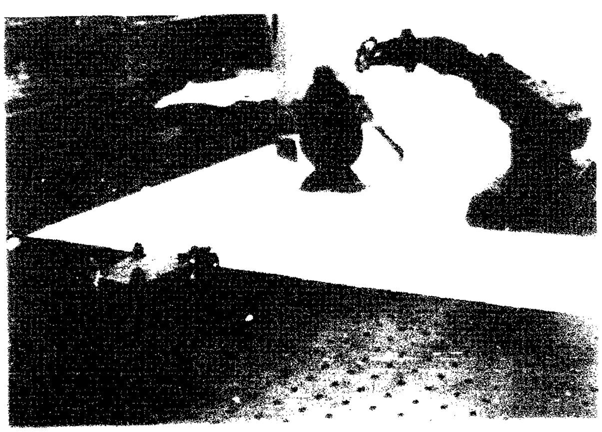
Then, assuming two frames as before, we obtain:

$$Q(Z_{22}, V_c, 1 - C) = \frac{V_c(1 - c)}{Z_{22}} = b_{21} - b_{22}$$

where for $i, j = 1, 2$, $b_{ij} = \dfrac{u_{nij}}{x_{ij}n_{xij} + y_{ij}n_{yij}}$

We thus see that we can compute at least the quantity $\dfrac{V_c(1 - c)}{Z}$, where Z is the depth at a point $p = (x, y)$ and where the direction of the normal flow (n_x, n_y) is perpendicular to the direction (T_x, T_y) of parallel translation, when the motion of the object is measured with regard to a coordinate system with origin at the object point whose image is point p and axes parallel to those of the camera coordinate system. Using the technique described in the Appendix, we can find the direction of parallel motion (T_x, T_y) for any position of the object coordinate system and choose that position for which the direction of the normal flow is





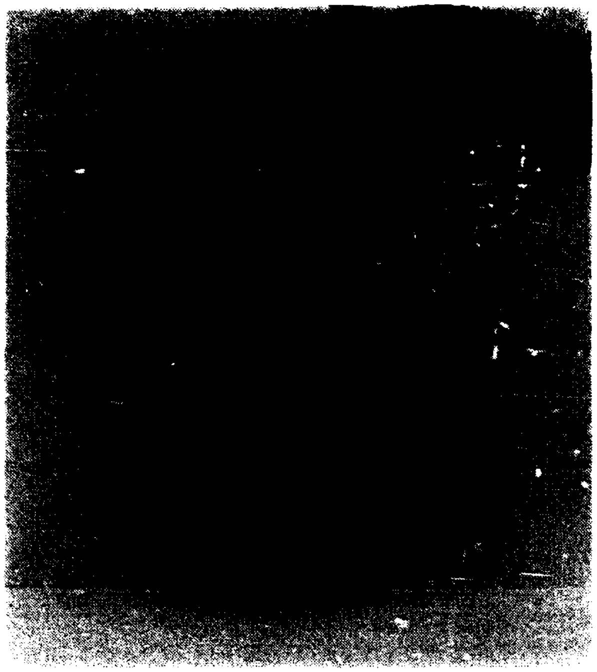


Figure 10 Normal flow of a moving robot arm with a stationary camera

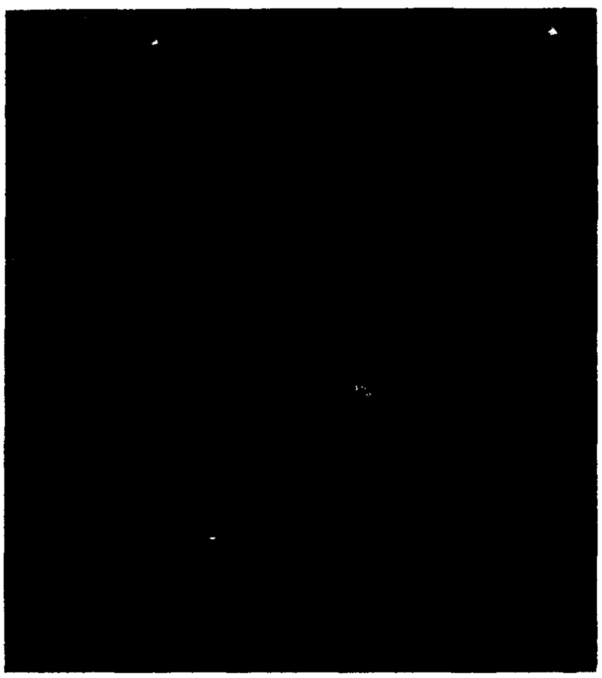


Figure 11 Image taken after both the robot arm and the camera have moved to new positions

$(Z/V_c(1 - c))$ was 10.230856 for the arm and 10.145772 for the toy, which agrees again with the ground truth.

These experiments demonstrate that the constraints introduced have the potential of giving rise to algorithms that can be used for the robust estimation of relative depth. Naturally, several stability issues need to

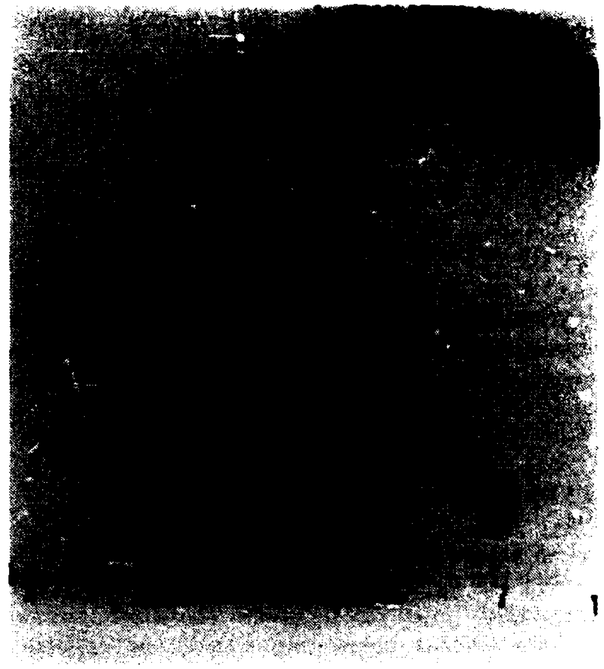


Figure 12 Normal flow of a moving robot arm and a moving camera

be examined. It is well known that particular motions of a visual sensor are quite pathological regarding the recovery of structure, while others are more stable. Such geometric facts need to be taken into account when we design active vision techniques and provide the sensor with an activity. In this particular case, the forward motion of the sensor might not be optimal, in the sense that it might not minimize errors in the estimation of relative depth.

SUMMARY AND CONCLUSIONS

We have presented a set of constraints relating relative depth and normal flow, i.e. the projection of the optic flow on the direction of the local intensity gradient, which we showed to be equal to the normal motion field in areas where the magnitude of the intensity gradient is large. The heart of the constraints lies in factoring out the effects of the parallel translation on the normal flow, by making measurements only at places where the normal flow is perpendicular to the parallel translation. Clearly, if nature conspired against this computational theory, it could present it with stimuli having only one or a few orientations, thus making it impossible to find normal flows perpendicular to the direction of parallel translation. However, for most objects in natural environments one can find gradients in almost any direction, and we should note that most moving objects have outlines which provide a (usually large) number of gradient directions. It is important to realize, however, that the procedures described here will never output an incorrect result. However, they may not be able to produce a result at all, in which case some other process should be used.

For the case of general translation we showed that relative depth can be computed at all points where the intensity gradient is perpendicular to the direction of the parallel translation. For the case of general motion we considered a coordinate system attached to any visible object point. The consequence of this is that at the image of that point the effect of the rotation on the normal flow is zero, and the solution proceeds as before, through the employment of a specific activity (acceleration along the optical axis). Clearly, many such points could be found.

APPENDIX

Here we describe a technique for finding the direction of parallel translation (V_x, V_y) from image measurements. We treat the problem in the general case (translation plus rotation). This appendix is a short summary of a technique described elsewhere[29]. In addition, we assume that the observer is 'looking' at the moving object, i.e. the object lies on the observer's optical axis. If this is not the case, the observer can always achieve it with a rotation of the camera (saccade). (It is important to realize, however, that such a saccade does not actually have to be implemented – it can be simulated, since the effects of a rotation are independent of depth. It is, of course, assumed here that the detection of the moving object has been accomplished[35,36].

Such a rotation introduces a known contribution to the normal flow. So, we assume that the moving object lies on the optical axis (*Figure A1*). To describe the motion of the object, we consider a coordinate system attached to it at its point of intersection with the optical axis. As a result, near the image origin the effect of rotation is negligible. Thus, considering a small area around the origin, we expect to find normal flows due to translation only. If we consider for simplicity a closed contour in that area (in an actual implementation one would have to consider all points inside the contour), then there are two possibilities for the pattern of normal flow (assuming that the object is moving closer). (If the object is moving away, the situation is symmetric.) It will be either as in *Figure A2* or otherwise (as, for example, in *Figures A3* and *A4*). If the pattern is as in *Figure A2*, then the FOE $\left(\frac{V_x}{V_z}, \frac{V_y}{V_z}\right)$ lies inside the contour and thus it is very small (negligible). Indeed, the FOE lies on the other side of the normal flow (*Figure A5*)[34].

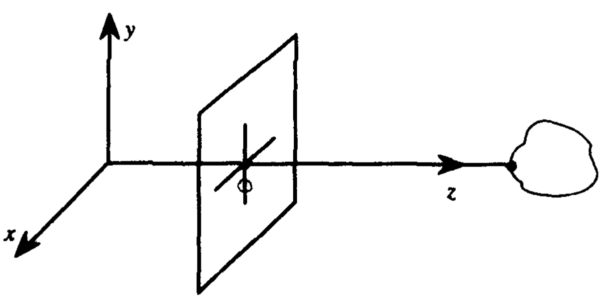


Figure A1 In actuality, not all lines will pass through the same point. In such a case, angle *AOB* gives all possible directions. Stability can be achieved if the analysis is done in the dual space, where each line corresponds to a point and a pencil of lines corresponds to a set of collinear points

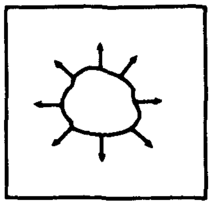


Figure A2

We need the direction of the vector (V_x, V_y). In fact, in our equations we only had vectors of the form $\left(\frac{V_x}{V_z}, \frac{V_y}{V_z}\right)$, which have the same direction as (V_x, V_y) (see *Figure A6*). But since $\left(\frac{V_x}{V_z}, \frac{V_y}{V_z}\right)$ has very small magnitude, the effect is the same, i.e. the quantity $\frac{V_x}{V_z}n_x + \frac{V_y}{V_z}n_y$ becomes negligible.

If the pattern is not as in *Figure A2*, there exists a dominant direction of the flow on the image plane. Assuming that the values of the flow are equal inside the small patch, we can compute the value of the flow from the normal flow values (*Figures A7* and *A8*). The direction of the flow at the origin is equal to the direction of parallel translation. Indeed, if (u, v) is the flow at the origin, we have $u = \frac{V_x}{Z}$, $v = \frac{V_y}{Z}$ where Z is the depth of the object point projecting to the origin. Thus $\frac{v}{u} = \frac{V_y}{V_x}$.

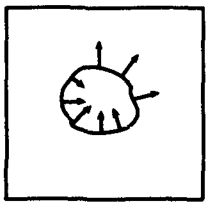


Figure A3

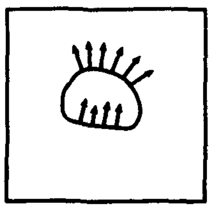


Figure A4

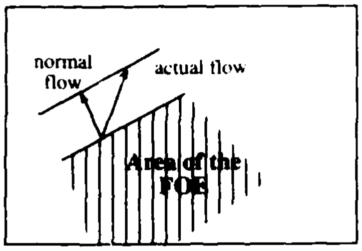


Figure A5

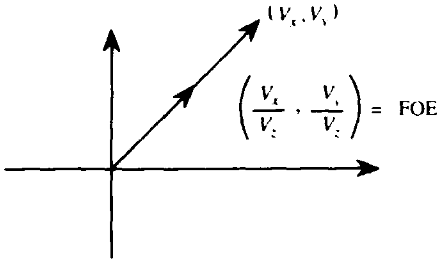


Figure A6

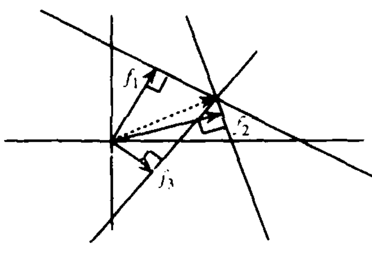


Figure A7

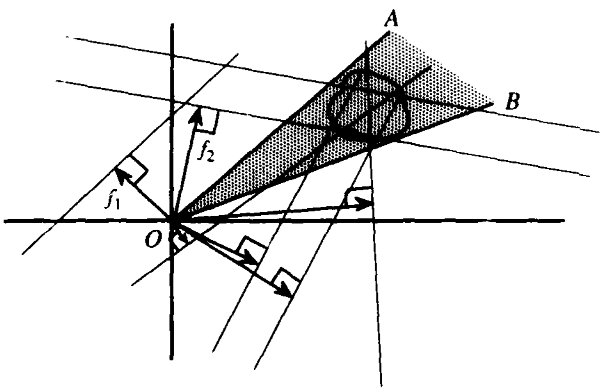


Figure A8

ACKNOWLEDGEMENTS

The support of ARPA, ONR and NSF (under a Presidential Young Investigator Award) is gratefully acknowledged.

REFERENCES

1 Bruss, A and Horn, B K P 'Passive navigation' *Comput. Vision Graph. & Image Process.* Vol 21 (1983) pp 3–21
2 Longuet-Higgins, H C 'A computer algorithm for reconstructing a scene from two projections', *Nature* Vol 293 (1981) pp 133–135
3 Longuet-Higgins, H C and Prazdny, K 'The interpretation of a moving retinal image', *Proc. Roy. Soc. Lond. B* Vol 208 (1984) pp 385–397
4 Ullman, S *The Interpretation of Visual Motion* MIT Press, Cambridge, MA (1979)
5 Spetsakis, M E and Aloimonos, Y 'Structure from motion using line correspondences', *Int. J. Comput. Vision* Vol 4 (1990) pp 171–183
6 Tsai, R Y and Huang, T S 'Uniqueness and estimation of three dimensional motion parameters of rigid objects with curved surfaces', *IEEE Trans. PAMI* Vol 6 (1984) pp 13–27
7 Waxman, A and Sinha, S S *Dynamic stereo: passive ranging to moving objects from relative image flows* Technical Report CAR-TR-74, Computer Vision Laboratory, Center for Automation Research, University of Maryland (1984)
8 Dickmanns, E D and Zapp, A 'A curvature based scheme for moving road vehicle guidance', *SPIE Proc. Conf. on Mobile Robots* Boston, MA (1986) pp 161–168
9 Dickmanns, E D '4D dynamic scene analysis with integral spatio-temporal models', in R Bolles and B Roth (eds.), *Robotics Research: 4th International Symposium*, MIT Press, Cambridge, MA (1988) pp 311–318
10 Harris, C G, Ibison, M C, Sparkes, E P and Stephens, M 'Structure and motion from optical flow', *Proc. 2nd Alvey Vision Conf.*, Bristol, UK (1986)
11 Harris, C G 'Determining ego-motion from matched points', *Proc. 3rd Alvey Vision Conf.* Cambridge, UK (September 1987) pp 189–192
12 Maybank, S J and Faugeras, O D 'Ambiguity in reconstruction from image correspondences', in O D Faugeras (ed.), *Proc. 1st Euro. Conf. on Comput. Vision* Antibes, France (1990) pp 177–186
13 Maybank, S J 'A second look at the least-squares algorithm for recovering information from optical flow', *Proc. 3rd Alvey Vision Conf.*, Cambridge, UK (September 1987) pp 221–226
14 Tistarelli, M and Sandini, G 'On the estimation of depth from motion using an anthropomorphic visual sensor', in O D Faugeras (ed.), *Proc. 1st Euro. Conf. on Comput. Vision*, Antibes, France (1990) pp 211–225
15 Marr, D *Vision*, W H Freeman, San Francisco, CA (1982)
16 Horn, B K P *Robot Vision*, McGraw-Hill, New York (1986)
17 Spetsakis, M E and Aloimonos, Y 'Optimal computing of structure from motion using point correspondences in two frames', *Proc. Int. Conf. on Comput. Vision*, Tarpon Springs, FL (1988) pp 449–453
18 Horn, B K P *Relative orientation*, AI Memo 994, MIT AI Laboratory, MA (1988)
19 Murray, D W and Buxton, B F *Experiments in the Machine Interpretation of Visual Motion*, MIT Press, Cambridge, MA (1990)
20 Aloimonos, A, Weiss, I and Bandopadhay, A 'Active vision', *Int. J. Comput. Vision*, Vol 2 (1988) pp 333–356
21 Horn, B K P and Schunck, B G 'Determining optical flow', *Artif. Intell.*, Vol 17 (1981) pp 185–203
22 Verri, A and Poggio, T 'Against quantitative optic flow', *Proc. Int. Conf. on Comput. Vision* (1987) pp 171–180
23 Adiv, G 'Determining three-dimensional motion and structure from optical flow generated by several moving objects', *IEEE Trans. PAMI*, Vol 7 (1985) pp 384–401
24 Adiv, G 'Inherent ambiguities in recovering 3d motion and structure from a noisy flow field', *Proc. Conf. on Comput. Vision and Patt. Recogn.*, Urbana, IL (1985) pp 70–77
25 Aloimonos, Y and Brown, C M 'Direct processing of curvilinear sensor motion from a sequence of perspective images', *Proc. Workshop on Comput. Vision: Representation and Control*, Annapolis, MD (1984) pp 72–77
26 Horn, B K P and Weldon Jr., E J 'Computationally-efficient methods of recovering translational motion', *Proc. Int. Conf. on Comput. Vision*, London, UK (1987) pp 1–11
27 Negahdaripour, S PhD thesis Technical report, MIT AI Laboratory, Cambridge, MA (1986)
28 Fermüller, C *Basic Visual Capabilities*, PhD thesis, Technical